

# Quantum broadcast channels

Jon Yard\* Patrick Hayden† Igor Devetak†

**Abstract**— We analyze quantum broadcast channels, which are quantum channels with a single sender and many receivers. Focusing on channels with two receivers for simplicity, we generalize a number of results from the network Shannon theory literature which give the rates at which two senders can receive a common message, while a personalized one is sent to one of them. Our first collection of results applies to channels with a classical input and quantum outputs. The second class of theorems we prove concern sending a common classical message over a quantum broadcast channel, while sending quantum information to one of the receivers. The third group of results we obtain concern communication over an isometry, giving the rates at quantum information can be sent to one receiver, while common quantum information is sent to both, in the sense that tripartite GHZ entanglement is established. For each scenario, we provide an additivity proof for an appropriate class of channels, yielding single-letter characterizations of the appropriate regions. We conclude with applications of the recently discovered state merging primitive, obtaining achievable rates for distributing independent quantum information among the parties in various ways, both with and without the assistance additional classical discussion among the receivers.

IN classical information theory, a discrete memoryless broadcast channel with a single sender Alice and two receivers, Bob and Charlie, is modeled by a probability transission matrix  $p(y, z|x)$ . The study of such channels was initiated by Cover in [4], where the idea of superimposing information in order to achieve rates of communication better than those achievable by naive timesharing protocols was introduced. There, it was also conjectured that the capacity region for sending common information at rate  $R$ , as well as independent personal information to each receiver at rates  $R_Y$  and  $R_Z$ , over a *degraded broadcast channel* (see e.g. [5] for a definition) consists of those triples of nonnegative rates  $(R, R_Y, R_Z)$  satisfying

$$\begin{aligned} R_Y &\leq I(X; Y|T) \\ R + R_Z &\leq I(T; Z) \end{aligned} \quad (1)$$

for some  $p(t, x)$ , where  $|T| \leq \min\{|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|\}$ . Cover's conjecture was validated by a coding theorem of Bergmans [3] together with a particularly clever proof of the single-letter converse by Gallager [15].

Such a result, however, gives no guarantee that the personal information being sent to each receiver will only be understandable at that receiver. One can require that *only* the intended receivers will be able to understand their own private messages. This problem was intially addressed by Wyner [31]

in the context of *wiretap channels*. Due to a well-established [8] operational correspondence between privacy and quantum coherence, a result of particular relevance to ours is one by Csiszar and Körner [7] which shows that *private* information can be sent to Bob at rate  $R_Y$ , while public information is sent to both Bob and Charlie at rate  $R$ , over an arbitrary broadcast channel  $p(y, z|x)$  if and only if there exists  $p(t, v)p(x|v)$  such that

$$\begin{aligned} 0 &\leq R_Y \leq I(V; Y|T) - I(V; Z|T) \\ 0 &\leq R \leq \min\{I(T; Y), I(T; Z)\}. \end{aligned} \quad (2)$$

In this paper, we consider quantum generalizations of the above problems in various settings, as we did in a similar manner for multiple access channels in [32]. We begin by considering common and personalized messages over channels with a classical input and quantum outputs. We then analyze the capabilities of arbitrary quantum broadcast channels for sending a common classical message to Bob and Charlie, while also sending quantum information to Bob, in the sense of *generating EPR entanglement* [8]. Our bounds on the quantum rates should be compared to those of Csiszar and Körner for sending private information to Bob, due to the privacy-coherence correspondence. Next, we show that for isometric channels, the common classical message can be made coherent, enabling the generation of GHZ entanglement among the three participants. Finally, we establish achievable regions for certain variations of the previous scenario in which all parties may communicate classically with each other for free in order to obtain various quantum correlations among themselves, providing applications of the recently discovered state merging primitive [21] for quantum information.

## I. PRELIMINARIES

### A. Classical and quantum systems

Throughout this paper, we use labels such as  $A, B, C$  to refer to quantum systems, writing  $\mathcal{H}_A$  for the Hilbert space whose unit vectors correspond to the pure states of the quantum system  $A$ . All Hilbert spaces will be finite dimensional, and we abbreviate  $\dim \mathcal{H}_A$  as  $|A|$ , so that  $\mathcal{H}_A \equiv \mathbb{C}^{|A|}$ . Given two systems  $A$  and  $B$ , the pure states of their composite system  $AB$  correspond to unit vectors in  $\mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$ . When we introduce a pure state, we use a superscripted label to identify the system to which the state refers. For example,  $|\phi\rangle^A \in \mathcal{H}_A$  and  $|\psi\rangle^{AB} \in \mathcal{H}_{AB}$ . The same convention will be followed when the state of a quantum system  $A$  is described by a density matrix, so that  $\rho^A \in \mathbb{C}^{|A| \times |A|}$  is a nonnegative definite Hermitian matrix with  $\text{Tr} \rho^A = 1$ . For a multipartite density matrix  $\rho^{ABC}$ , we frequently abbreviate its partial traces as  $\rho^{AB} = \text{Tr}_C \rho^{ABC}$ . In later references to the global

\* jtyard@cs.caltech.edu, Institute for Quantum Information, California Institute of Technology, Pasadena, California, USA

† patrick@cs.mcgill.ca, School of Computer Science, McGill University, Montréal, Canada

† devetak@usc.edu, Electrical Engineering Department, University of Southern California, USA

state, we frequently drop the superscript completely, although the partial traces will always have superscripts. We often use the abbreviation  $\phi \equiv |\phi\rangle\langle\phi|$  when referring to the rank-one density matrix corresponding to a pure state  $|\phi\rangle$ .

For an arbitrary matrix  $M \in \mathbb{C}^{d \times d}$ , its *trace norm*  $|M|_1$  is defined as the sum of its singular values, expressed as  $|M|_1 = \text{Tr} \sqrt{MM^\dagger}$ . Given two states  $\rho^A$  and  $\sigma^A$ , their *trace distance*  $|\rho - \sigma|_1$  is the trace norm of their difference. We use the squared version of the *fidelity*, defined as  $F(\rho, \sigma) = |\sqrt{\rho}\sqrt{\sigma}|_1^2$ . When  $\rho = |\phi\rangle\langle\phi|$ , the fidelity evaluates to  $F(|\phi\rangle\langle\phi|, \sigma) = \langle\phi|\sigma|\phi\rangle$ . These distances are related [14] via

$$F(\rho, \sigma) \geq 1 - |\rho - \sigma|_1 \quad (3)$$

$$|\rho - \sigma|_1 \leq 2\sqrt{1 - F(\rho, \sigma)}. \quad (4)$$

Since the trace distance comes from a norm, it satisfies the usual triangle inequality

$$|\rho_1 - \rho_3|_1 \leq |\rho_1 - \rho_2|_1 + |\rho_2 - \rho_3|_1.$$

We shall frequently make use of classical-quantum states and classical-quantum channels [10] in this paper. To any finite set  $\mathcal{X}$ , we associate a Hilbert space  $\mathcal{H}_X$  with orthonormal basis  $\{|x\rangle^X\}_{x \in \mathcal{X}}$ , so that for any classical random variable  $X$  which takes the value  $x \in \mathcal{X}$  with probability  $p(x)$ , we may write a density matrix

$$\rho^X = \sum_x p(x) |x\rangle\langle x|^X \equiv \bigoplus_x p(x)$$

which is diagonal in that basis. For any  $\mathcal{S} \subseteq \mathcal{X}$ , if  $P_{\mathcal{S}}$  for the projector onto the subspace spanned by  $\{|x\rangle^X\}_{x \in \mathcal{S}}$ , we then have

$$\text{Pr}\{X \in \mathcal{S}\} = \text{Tr} P_{\mathcal{S}} \rho^X = \sum_{x \in \mathcal{S}} p(x).$$

An ensemble of quantum states  $\{\rho_x^B, p(x)\}$  can be represented in a similar way with a block diagonal *classical-quantum (cq) state*

$$\rho^{XB} = \sum_x p(x) |x\rangle\langle x|^X \otimes \rho_x^B \equiv \bigoplus_x p(x) \rho_x^B.$$

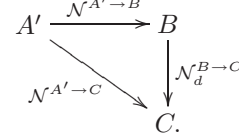
Wherever possible, we will adopt the more compact direct sum notation  $\bigoplus$  for describing cq states, with the understanding that the labels of the blocks correspond to states of an additional classical system.

## B. Channels

A *classical-quantum (cq) channel*  $W^{\mathcal{X} \rightarrow B}$  describes a physical setup in which the sender Alice is able to remotely prepare any one of a collection of *conditional density matrices*  $\{\rho_x^B\}_{x \in \mathcal{X}}$  in the laboratory of Bob. By a *cq broadcast channel*  $W^{\mathcal{X} \rightarrow BC}$  from Alice to Bob and Charlie, we mean a physical scenario in which Alice prepares any one of a collection of bipartite conditional density matrices  $\{\rho_x^{BC}\}_{x \in \mathcal{X}}$ .

By a *quantum channel*  $\mathcal{N}^{A' \rightarrow B}$  from  $A'$  to  $B$ , we mean a trace-preserving linear map from density matrices on  $A'$  to those on  $B$  which is also *completely positive*. Such an operator may be referred to as a *map* in this text, usually when it does not represent a physical process between spatially separated parties. Here, we parallel the state convention by treating the

superscript  $A' \rightarrow B$  as a definition of the domain and range of the channel, to be omitted in later references to  $\mathcal{N}$ . In this paper, a *quantum broadcast channel*  $\mathcal{N}^{A' \rightarrow BC}$  refers to a quantum channel with a single input and two outputs. We often personify the users of the channel, saying that Alice controls the input, while Bob and Charlie are located at the respective outputs. Defining the channel from Alice to Bob as  $\mathcal{N}^{A' \rightarrow B} \equiv \text{Tr}_C \mathcal{N}^{A' \rightarrow BC}$ , with a similar definition for  $\mathcal{N}^{A' \rightarrow C}$ , we will say that the broadcast channel  $\mathcal{N}^{A \rightarrow BC}$  is *degraded* whenever there exists a *degrading channel*  $\mathcal{N}_d^{B \rightarrow C}$  from Bob to Charlie satisfying  $\mathcal{N}^{A' \rightarrow C} = \mathcal{N}_d^{B \rightarrow C} \circ \mathcal{N}^{A' \rightarrow B}$ . In other words, the following diagram must commute:



*Remark 1:* In the classical literature, such channels have been called *stochastically degraded*, meaning that the random variables  $X, Y$  and  $Z$ , analogous to  $A, B$  and  $C$  of the state  $\rho^{ABC} = \mathcal{N}^{A' \rightarrow BC}(\phi^{AA'})$ , form a Markov chain  $X - Y - Z$ . However, in a quantum Markov chain [18]  $A - B - C$  with state  $\rho^{ABC}$ , there must exist a recovery map  $\mathcal{M}^{B \rightarrow BC}$  satisfying  $\mathcal{M}(\rho^{AB}) = \rho^{ABC}$ . In our case we have the weaker condition  $\mathcal{N}_d^{B \rightarrow C}(\rho^{AB}) = \rho^{AC}$ . The two conditions are equivalent in the classical problem because classical information can be copied.

Given a channel  $\mathcal{N}^{A' \rightarrow B}$ , there always exists an isometry  $\mathcal{U}^{A' \rightarrow BE}$  into an unobservable environment which *extends* the channel, meaning that  $\mathcal{N}^{A' \rightarrow B} = \text{Tr}_E \mathcal{U}^{A' \rightarrow BE}$ . We will call such an isometry an *isometric extension* of  $\mathcal{N}^{A' \rightarrow B}$ . While there are generally many choices for an isometric extension of a given channel, all are related via isometries on the environment  $E$ . On the other hand, any channel obtained by disregarding the output  $B$  of such an isometric extension will be said to be *complementary* to  $\mathcal{N}^{A' \rightarrow B}$ , which we write  $\mathcal{N}_c^{A' \rightarrow E} = \text{Tr}_B \mathcal{U}^{A' \rightarrow BE}$ . In case the isometric extension  $\mathcal{U}^{A' \rightarrow BE}$  of  $\mathcal{N}^{A' \rightarrow B}$  is a degraded broadcast channel, in the sense that there is a degrading map  $\mathcal{N}_d^{B \rightarrow E}$  for which  $\mathcal{N}_c^{A' \rightarrow E} = \mathcal{N}_d^{B \rightarrow E} \circ \mathcal{N}^{A' \rightarrow B}$ , we will say that the channel  $\mathcal{N}_c^{A' \rightarrow E}$  is *degradable* [9]. Concrete examples of degradable channels include erasure channels [2], qubit flip channels, and photon number splitting channels [16].

A particular class of degradable channels which are relevant to this paper are the *generalized dephasing channels* [9], [32]. These are channels  $\mathcal{N}^{A' \rightarrow B}$  with  $|A| = |B|$  which act noiselessly on some common orthonormal basis  $\{|x\rangle^A, |x\rangle^B\}$ . Such channels have an isometric extension

$$\mathcal{U}^{A' \rightarrow BE} = \sum_x |\phi_x\rangle^E |x\rangle^B \langle x|^A$$

for some (not necessarily orthogonal) normalized vectors  $\{|\phi_x\rangle^E\}$ , and a complementary channel acting as

$$\mathcal{N}_c(\rho) = \sum_x \langle x|\rho|x\rangle \phi_x^E.$$

Writing

$$\Delta^{A' \rightarrow B}: \rho \mapsto \sum_x |x\rangle\langle x|\rho|x\rangle\langle x| \quad (5)$$

for the *completely dephasing channel*, which sets to zero all off-diagonal matrix elements, any generalized dephasing channel  $\mathcal{N}^{A' \rightarrow B}$  satisfies

$$\mathcal{N}_c \circ \Delta = \mathcal{N}_c \quad (6)$$

$$H(\Delta(\rho)) \geq H(\mathcal{N}(\rho)). \quad (7)$$

For the decoding of classical information, we use (somewhat interchangeably) the notions of POVM's and *quantum instruments*  $\mathcal{D}^{A \rightarrow BX}$ . The latter is a quantum channel whose target is a cq system. Such a map can be specified in terms of a collection of (generally) trace-reducing maps  $\{\mathcal{D}_x^{A \rightarrow B}\}$  for which  $\sum_x \mathcal{D}_x$  is trace-preserving. The instrument then acts as  $\mathcal{D}(\rho^A) = \bigoplus_x \mathcal{D}_x(\rho^A)$ . Given a POVM  $\{\Lambda_x\}$  on  $A$ , its *associated measurement instrument*  $\mathcal{D}^{A \rightarrow X}$  has components acting as  $\mathcal{D}_x(\rho^A) = \text{Tr } \Lambda_x \rho$ .

### C. Entropy and information quantities

Let  $\rho^{ABC}$  be any tripartite density matrix. We write  $H(A)_\rho = H(\rho^A) \equiv -\text{Tr}(\rho^A \log \rho^A)$  for the *von Neumann entropy* of the reduced density matrix  $\rho^A$ , omitting the subscripted state when it is apparent. As is common with much of quantum Shannon theory, certain linear combinations of entropies of various subsystems of the joint state  $\rho^{ABC}$  arise naturally in the characterizations of the various rate regions we will introduce. We review the essential ones here, beginning with the *conditional entropy*

$$H(A|B) = H(AB) - H(B).$$

This quantity is defined in direct analogy to its counterpart in classical information theory, which is always positive and can be regarded as an average entropy of conditional probability distributions. It is however clear that  $H(A|B) = -1$  when evaluated on an EPR state  $\frac{1}{\sqrt{2}}(|00\rangle^{AB} + |11\rangle^{AB})$ , a fact which was long considered problematic for the use conditional entropy in quantum information theory. Nonetheless, the negative of conditional entropy has been defined as the *coherent information*

$$I(A) \rangle B = -H(A|B)$$

from  $A$  to  $B$ , due to its utility in characterizing the capacity of a quantum channel for transmitting quantum information [24], [28], [8] as a certain optimization problem which always yields a nonnegative rate. Following [26], we may also write

$$I_c(\rho^{A'}, \mathcal{N}^{A' \rightarrow B}) \equiv I(A) \rangle B_{\mathcal{N}(\varphi)},$$

where  $|\varphi\rangle^{AA'}$  is any purification of  $\rho^{A'}$ . An operational interpretation of both positive and negative conditional entropies was recently given in [21], where the primitive of *state merging* was introduced, yielding a quantum counterpart to the classical Slepian-Wolf theorem for distributed data compression. We will use this merging primitive in Section II-D to transmit quantum information over a broadcast channel.

The mutual information and conditional mutual information are respectively defined as

$$I(A; B) = H(A) - H(A|B) = H(A) + H(B) - H(AB)$$

and as

$$\begin{aligned} I(A; B|C) &= H(A|C) - H(A|BC) \\ &= I(A; BC) - I(A; C) \\ &= I(AC; B) - I(B; C). \end{aligned} \quad (8)$$

By the *strong subadditivity* [23] of quantum entropy, it follows that mutual information and conditional mutual information are nonnegative. There are many equivalent formulations of strong subadditivity which we will now recall. By simple algebra,  $I(A; B|C) \geq 0$  is seen to be equivalent to the inequality  $H(A|BC) \leq H(A|B)$  which is interpreted as saying that *conditioning reduces entropy*, and thus increases coherent information  $I(A) \rangle BC \geq I(A) \rangle B$ . These can easily be used to derive either form of the *data processing inequality*, which say that given any channel  $\mathcal{N}^{B \rightarrow C}$ ,

$$I(A; B)_{\rho^{AB}} \geq I(A; C)_{\mathcal{N}(\rho^{AB})} \quad (9)$$

$$I(A) \rangle B_{\rho^{AB}} \geq I(A) \rangle C_{\mathcal{N}(\rho^{AB})}. \quad (10)$$

In other words, processing the output of a channel will never increase the mutual or coherent information over that channel. We remark that the first inequality above includes the Holevo bound [19] as a special case, since a measurement can be considered as a quantum channel with a strictly classical output. Note that (10) can also be written

$$I_c(\rho^{A'}, \mathcal{N}^{A' \rightarrow B}) \leq I_c(\rho^{A'}, \mathcal{M}^{B \rightarrow C} \circ \mathcal{N}^{A' \rightarrow B})$$

for every  $\rho^{A'}$ ,  $\mathcal{N}^{A' \rightarrow B}$  and  $\mathcal{M}^{B \rightarrow C}$ . Finally, given a quadripartite system  $A_1 A_2 B_1 B_2$ , the following inequality is implied by and also implies strong subadditivity:

$$H(A_1 A_2 | B_1 B_2) \leq H(A_1 | B_1) + H(A_2 | B_2). \quad (11)$$

## II. MAIN RESULTS

### A. Degraded message sets for cq channels

In what follows, a sequence  $x_1 x_2 \dots x_n$ , with each  $x_i$  belonging to some set  $\mathcal{X}$  will be denoted by  $x^n$ . Using many instances of a cq broadcast channel  $W^{\mathcal{X} \rightarrow BC}$ , suppose that Alice wishes send a personal message to Bob while simultaneously sending an independent common message to Bob and Charlie. If  $W$  has conditional density matrices  $\rho_x^{BC}$ , we define an  $(R, R_B, n, \epsilon)$  code for  $W$  to consist of an encoding  $\{x^n(m, k) \in \mathcal{X}^n\}$  where  $(m, k) \in 2^{nR} \times 2^{nR_B}$ , a POVM  $\{\Lambda_{mk}\}$  on  $B^n$  and a POVM  $\{\Lambda'_m\}$  on  $C^n$  which satisfy

$$\text{Tr } \rho_{x^n(m, k)} (\Lambda_{mk} \otimes \Lambda'_m) \geq 1 - \epsilon$$

for every  $(m, k) \in 2^{nR} \times 2^{nR_B}$ . A rate pair  $(R, R_B)$  is *achievable* if there is a sequence of  $(R, R_B, n, \epsilon_n)$  codes with  $\epsilon_n \rightarrow 0$ . The *classical capacity region*  $\mathcal{C}(W)$  of  $W$  is defined as the closure of the collection of all such achievable rate triples.

In Theorem 1, we give a regularized expression for  $\mathcal{C}(W)$ , generalizing a coding theorem from [3], though we also prove a multi-letter converse. Theorem 2 shows that for a class of degraded cq channels, that characterization can be single-letterized, generalizing the classical converse of [15].



*Theorem 1:* Given a cq channel  $W^{\mathcal{X} \rightarrow BC}$  with conditional density matrices  $\{\rho_x^{BC}\}$ ,  $\mathcal{C}(W)$  equals the closure of the pairs of nonnegative rates  $(R, R_B)$  which satisfy

$$\begin{aligned} R_B &< I(X^k; B^k|T)_\sigma/k \\ R &< \min\{I(T; B^k)_\sigma, I(T; C^k)_\sigma\}/k \end{aligned}$$

for some  $k \geq 1$  and some  $p(t, x^k)$  giving rise to

$$\sigma^{TX^k B^k C^k} = \bigoplus_{t, x^k} p(t, x^k) \rho_{x^k}^{B^k C^k} \quad (12)$$

with  $|T| \leq \min\{|\mathcal{X}|^k, |B|^{2k} + |C|^{2k} - 1\}$ . Here,  $\rho_{x^k} = \bigotimes_k \rho_{x_k}$ .

*Theorem 2:* Suppose that the conditional density matrices  $\{\rho_x^{BC}\}$  of a cq channel  $W^{\mathcal{X} \rightarrow BC}$  are such that their restrictions  $\{\rho_x^B\}$  mutually commute, and that the restrictions  $\{\rho_x^C\}$  satisfy  $\rho_x^C = \mathcal{M}(\rho_x^B)$  for some channel  $\mathcal{M}^{B \rightarrow C}$ . Then

$$\begin{aligned} R_B &\leq I(X; B|T)_\sigma \\ R &\leq I(T; C)_\sigma \end{aligned}$$

for some state  $\sigma^{TXBC}$  of the form (12) with  $k = 1$  and  $|T| \leq \min\{|\mathcal{X}|, |B|^2\}$ .

### B. Classical-quantum region $\mathcal{CQ}(\mathcal{N})$ for quantum channels

In this scenario, Alice wishes to send quantum information to Bob at rate  $Q$ , while sending common classical information to Bob and Charlie. To this end, she prepares one of many states  $\{|\Upsilon_m\rangle^{AA'^n}\}$  which are entangled between a system  $A$  in her laboratory and the inputs of some large number of parallel identical broadcast channels. Bob employs a quantum instrument  $\mathcal{D}_1^{B^n \rightarrow \hat{A} M_B}$ , with the goal of learning the classical message, as well as holding the  $\hat{A}$  part of a highly entangled state. Meanwhile, Charlie performs a measurement, modeled by the instrument  $\mathcal{D}_2^{C^n \rightarrow \hat{M}_C}$ , to learn the common classical message. Such components will be said to comprise an  $(R, Q, n, \epsilon)$  cq entanglement generation code for the broadcast channel  $\mathcal{N}^{A' \rightarrow BC}$  if, for  $|m\rangle^{M_B M_C} \equiv |m\rangle^{M_B} |m\rangle^{M_C}$ ,

$$F(|m\rangle^{M_B M_C} | \Phi_Q \rangle^{A\hat{A}}, (\mathcal{D}_1 \otimes \mathcal{D}_2) \circ \mathcal{N}^{\otimes n}(\Upsilon_m^{AA'^n}) \geq 1 - \epsilon$$

for each  $m$ , where  $|\Phi_Q\rangle^{A\hat{A}}$  is some fixed rate  $Q$  EPR state

$$|\Phi_Q\rangle^{A\hat{A}} = \frac{1}{\sqrt{2^{nQ}}} \sum_{a=1}^{2^{nQ}} |a\rangle^A |a\rangle^{\hat{A}}.$$

A pair of nonnegative rates  $(Q, R)$  is called an *achievable cq rate pair for entanglement generation* if there exists a sequence of  $(Q, R, n, \epsilon_n)$  cq entanglement generation codes with  $\epsilon_n \rightarrow 0$ . The *cq capacity region for entanglement generation*  $\mathcal{CQ}(\mathcal{N})$  is defined as the closure of the set of such achievable cq rate pairs. The following theorem describes  $\mathcal{CQ}(\mathcal{N})$  of any broadcast channels as a regularized union of rectangles.

*Theorem 3:* Let  $\mathcal{N}^{A' \rightarrow BC}$  be arbitrary. Then  $\mathcal{CQ}(\mathcal{N})$  is equal to the closure of the collection of pairs of nonnegative cq rates  $(Q, R)$  satisfying

$$\begin{aligned} Q &< I(A)B^k T)_\sigma/k \\ R &< \min\{I(T; B^k)_\sigma, I(T; C^k)_\sigma\}/k \end{aligned}$$

for some  $k \geq 1$  and some state

$$\sigma^{TAB^k C^k} = \bigoplus_t p(t) \mathcal{N}^{\otimes k}(\phi_t^{AA'^k}). \quad (13)$$

arising from the action of  $\mathcal{N}^{\otimes k}$  on the  $A'^k$  parts of some bipartite pure state ensemble  $\{p(t), |\phi_t\rangle^{AA'^k}\}$ . Moreover, it suffices to take  $|T| \leq \min\{|A'|^{2k}, |B|^{2k} + |C|^{2k} - 1\}$ .

Our next theorem gives a single-letter characterization of  $\mathcal{CQ}$  whenever Charlie holds part of the environment of a generalized dephasing channel from Alice to Bob.

*Theorem 4:* Let  $\mathcal{N}^{A' \rightarrow BC}$  have an isometric extension

$$\mathcal{U} = \sum_x |x\rangle^B |\psi_x\rangle^{CE} \langle x|^{A'}$$

so that  $\mathcal{N}^{A' \rightarrow B}$  is a generalized dephasing channel. Then  $\mathcal{CQ}(\mathcal{N})$  equals those pairs of nonnegative cq rates  $(Q, R)$  satisfying

$$\begin{aligned} Q &\leq H(X|T) - H(CE|T) \\ R &\leq I(T; C) \end{aligned}$$

for some state

$$\omega^{TXCE} = \bigoplus_{t,x} p(t, x) \psi_x^{CE}$$

with  $|T| \leq |\mathcal{X}|$ .

In particular, this theorem applies to any isometric extension of the following *pinching channel*  $\mathcal{P}: \mathbb{C}^{3 \times 3} \rightarrow \mathbb{C}^{3 \times 3}$ , which acts by setting some matrix elements to zero, while leaving the others alone, according to

$$\mathcal{P}: \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \mapsto \begin{pmatrix} * & * & \\ * & * & \\ & & * \end{pmatrix},$$

Mathematically, this channel is a completely positive trace-preserving conditional expectation of  $\mathbb{C}^{3 \times 3}$  onto a \*-subalgebra which is spatially isomorphic to  $\mathbb{C}^{2 \times 2} \oplus \mathbb{C}$ . For the broadcast channel corresponding to any isometric extension  $\mathcal{U}_{\mathcal{P}}^{A' \rightarrow BC}$  of  $\mathcal{P}$  where Charlie obtains the entire environment of  $\mathcal{P}$ , a straightforward derivation reveals that the outer boundary of  $\mathcal{CQ}(\mathcal{U}_{\mathcal{P}})$  is given by

$$\begin{aligned} Q_B &= p \\ R &= \begin{cases} 1 & \text{if } p \leq 1/2 \\ H(p) & \text{if } p \geq 1/2 \end{cases} \end{aligned}$$

where  $0 \leq p \leq 1$ , as is shown in Figure II-B.

### C. Quantum region $\mathcal{Q}(\mathcal{N})$ for quantum channels

Here, Alice attempts to share a large bipartite entangled state with Bob, while also trying to build a large GHZ state with Bob and Charlie. Alice encodes by preparing the state  $|\Upsilon\rangle^{AGA'^n}$ , entangled with the inputs of a large number  $n$  of instances of  $\mathcal{N}^{A' \rightarrow BC}$ . Bob and Charlie employ respective decoding maps  $\mathcal{D}_1^{B^n \rightarrow \hat{A} G_B}$  and  $\mathcal{D}_2^{C^n \rightarrow G_C}$ . These components comprise a  $(Q, Q_B, n, \epsilon)$  entanglement generation code for

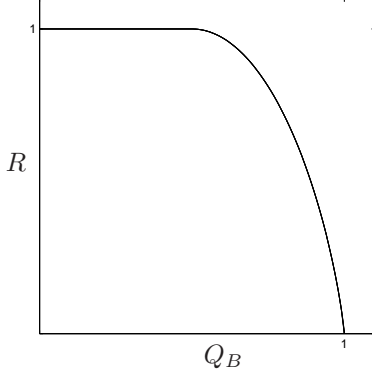


Fig. 1.  $\mathcal{CQ}$  for pinching channel

the broadcast channel  $\mathcal{N}$  if they generate a rate  $Q_B$  EPR state  $|\Phi_{Q_B}\rangle^{A\hat{A}}$  and a rate  $Q$  GHZ state

$$|\Gamma_Q\rangle^{GG_B G_C} = \frac{1}{\sqrt{2^{nQ}}} \sum_{m=1}^{2^{nQ}} |m\rangle^G |m\rangle^{G_B} |m\rangle^{G_C}$$

in the sense that

$$F\left(|\Phi_{Q_B}\rangle^{A\hat{A}} |\Gamma_Q\rangle^{GG_B G_C}, (\mathcal{D}_1 \otimes \mathcal{D}_2)(\Upsilon^{GAA^n})\right) \geq 1 - \epsilon.$$

Achievable rates and the capacity region  $\mathcal{Q}(\mathcal{N})$  are defined in analogy to the earlier scenarios. In Theorem 5, we give a multi-letter formula for  $\mathcal{Q}(\mathcal{N})$  in the case where  $\mathcal{N}$  is an isometry. Theorem 6 derives a single-letter formula for  $\mathcal{Q}(\mathcal{N})$  in case  $\mathcal{N}$  is an isometric extension of a generalized dephasing channel to Bob. Note that these results can be regarded as dynamic analogs of those obtained by [29], who study distillation of EPR and GHZ entanglement from arbitrary tripartite pure states. While those authors allow for additional classical communication, we do not. Similar correspondences exist in the literature, such as between [10] and [8], as well as between [21] and [32].

**Theorem 5:** Let  $\mathcal{U}^{A' \rightarrow BC}$  be an isometric broadcast channel. Then  $\mathcal{Q}(\mathcal{U})$  is given by the closure of the pairs of nonnegative quantum rates  $(Q, Q_B)$  satisfying

$$\begin{aligned} Q_B &< I(A)B^k T)_\sigma / k \\ Q &< \min\{I(T; B^k)_\sigma, I(T; C^k)_\sigma\} / k \end{aligned}$$

where  $\sigma^{TAB^k C^k}$  takes the same form as in (13), replacing  $\mathcal{N}$  with  $\mathcal{U}$ . The bound on  $|T|$  is the same as well.

**Theorem 6:** Let  $\mathcal{U}^{A' \rightarrow BC}$  be a broadcast channel which is an isometric extension of a generalized dephasing channel to  $B$ , written

$$\mathcal{U} = \sum_x |x\rangle^B |\psi_x\rangle^C \langle x|^{A'}.$$

Then  $\mathcal{Q}(\mathcal{U})$  equals the set of pairs of nonnegative quantum rates  $(Q_B, Q)$  satisfying

$$\begin{aligned} Q_B &\leq H(X|T)_\omega - H(C|T)_\omega \\ Q &\leq I(T; C)_\omega \end{aligned}$$

where

$$\omega^{TXC} = \bigoplus_{x,t} p(t, x) \psi_x^C$$

and  $|\mathcal{T}| \leq |\mathcal{X}|$ .

When  $\mathcal{U}^{A' \rightarrow BC}$  is an isometric extension of the pinching channel  $\mathcal{P}^{A' \rightarrow B}$ , this theorem yields the rate region from Figure II-B with  $R$  replaced by  $Q$ .

**Remark 2:** Using the standard technique of restricting to a high-fidelity subspace of the input, it is possible to strengthen the previous four theorems to obtain stronger error criteria, such as that from the strong subspace transmission of [32]. We have, however, focused on entanglement generation for simplicity.

#### D. Achievable quantum rates from state merging

Suppose that three players, Alice, Bob and Charlie, share the respective parts of many instances of a tripartite pure state  $|\psi\rangle^{ABC}$ . Assuming that Alice can send classical bits to Bob for free, it was recently shown [21] that the quantum communication cost for Alice to transfer her  $A^n$  systems to Bob is asymptotically equal to  $H(A|B)$ , regardless of the negativity of the expression. Specifically, whenever  $H(A|B)$  is negative (or equivalently, when  $I(A)B) > 0$ ), Alice and Bob can generate EPR entanglement at rate  $I(A)B)$  in the process of transferring  $A^n$  to Bob, using *only* classical communication and no quantum communication whatsoever. This is an improvement over standard entanglement distillation [11], where the same amount of EPR entanglement is obtained without deliberately trying to accomplish state merging. On the other hand, in case  $H(A|B) > 0$ , the protocol requires as input Alice-Bob EPR entanglement at a rate of at least  $H(A|B)$  ebits per system to be transferred. Therefore, if Alice and Bob performed state merging yesterday at a negative cost, they can use the extra entanglement they generated to perform merging with a positive cost today. As in [21], we only directly consider state merging when the cost is negative, as the protocol with positive cost is obtained by having Alice establish an appropriate amount of pure entanglement with Bob, so that the total coherent information they share becomes positive. Formally, a *negative cost state merging protocol* for a state  $|\psi\rangle^{ABC}$  consists of an instrument  $\mathcal{M}^{A^n \rightarrow DM}$  with components  $\mathcal{M}_m^{A^n \rightarrow D}$  to be performed on Alice's systems  $A^n$ , together with a collection of decoding operations  $\mathcal{D}_m^{B^n \rightarrow B^n \hat{A} \hat{D}}$  for Bob. The quantum outputs  $D$  and  $\hat{D}$  hold Alice's and Bob's respective halves of the entanglement resulting from the protocol, while each copy of  $\hat{A}$  corresponds to a system located in Bob's laboratory which is isomorphic to  $A$ , whose purpose is to hold the corresponding part of the transferred states. The protocol proceeds as follows. Alice performs the instrument  $\mathcal{M}^{A^n \rightarrow DM}$  and tells the classical result  $M$  to Bob who, depending on that classical data he receives, uses the appropriate decoding map. These components will be said to comprise a  $(Q, n, \epsilon)$  *negative cost state merging protocol* for  $|\psi\rangle^{ABC}$  if  $|D| = |\hat{D}| = 2^{nQ}$  and

$$F\left(|\psi\rangle^{\otimes n} |\Phi_Q\rangle^{D\hat{D}}, \sum_m (1^{C^n} \otimes \mathcal{D}_m \otimes \mathcal{M}_m)(\psi^{\otimes n})\right) \geq 1 - \epsilon,$$

where  $|\Phi_Q\rangle^{D\hat{D}}$  is a rate  $Q$  maximally entangled state. The following proposition is from [21], and is proved in [22].

*Proposition 1:* Let a pure tripartite state  $|\psi\rangle^{ABC}$  satisfying  $I(A)B > 0$  be given. Then, for every  $\epsilon > 0$ , and every  $0 \leq Q < I(A)B$ , there is  $n$  sufficiently large so that there exists a  $(Q, n, \epsilon)$  negative cost state merging protocol for  $|\psi\rangle^{ABC}$ . We now state a theorem.

*Theorem 7:* Let  $\mathcal{N}^{A' \rightarrow BC}$  be arbitrary. If Bob can communicate for free with Charlie via a classical channel, then Alice may generate rate  $Q_C$  entanglement with Charlie whenever there is a bipartite pure state  $|\psi\rangle^{AA'}$  for which

$$Q_C < I(A)BC)_\sigma \text{ and } I(B)C)_\sigma > 0,$$

where  $\sigma^{ABC} = \mathcal{N}(\psi^{AA'})$ . In addition, the same protocol allows Bob and Charlie to generate independent EPR entanglement between themselves at any rate less than  $I(B)C)$ .

Here, we give an outline of the proof:

*Proof:* Assume that Alice and Charlie share common randomness. Fixing a single-letter reference state  $|\psi\rangle^{AA'}$  satisfying the conditions of the theorem, Alice uses a random LSD code of rate  $I(A)BC)_\sigma$  (see Proposition 4) based on her common randomness with Charlie, pretending as though Bob and Charlie can collaborate in their decoding. As her average code density matrix is close to the product state  $(\psi^{A'})^{\otimes n}$ , the output state of Bob and Charlie is close to  $(\sigma^{BC})^{\otimes n}$ . By assumption,  $I(B)C)_\sigma > 0$ , so there is a *negative cost* for Bob to transfer his  $B^n$  systems to Charlie. This means that Bob and Charlie can distill EPR's at any rate less than  $I(B)C)_\sigma$  during this process. Charlie uses the common randomness to decode the random LSD code, thus establishing the rate  $I(A)BC)_\sigma$  entanglement with Alice. Finally, the protocol is derandomized using standard arguments. ■

Finally, we demonstrate that Alice may generate, and also *transmit* [1] independent entanglement between herself and each receiver *without* the assistance of classical communication between the two receivers.

*Theorem 8:* Let  $\mathcal{N}^{A' \rightarrow BC}$  be arbitrary, and let  $|\psi\rangle^{ABAC'A'}$  be entangled between local systems  $A_B$  and  $A_C$  in Alice's lab and the  $A'$  input to the channel. Provided that  $I(A_B)B)_{\mathcal{N}(\psi)}$  and  $I(A_C)C)_{\mathcal{N}(\psi)}$  are positive, Alice may generate those same amounts of independent entanglement with each receiver. We will be content to outline the proof:

*Proof:* If communication is allowed between Alice and each of the receivers, the theorem is immediate from a double application of Proposition 1 (or, rather, entanglement distillation [11], as the state merging aspect is not needed). We now argue that the classical communication is not needed. Alice begins such a protocol by applying instruments  $\mathcal{E}^{A_B \rightarrow \tilde{B}M}$  and  $\mathcal{E}'^{A_C \rightarrow \tilde{C}K}$  with components  $\{\mathcal{E}_m^{A_B \rightarrow \tilde{B}}\}_m$  and  $\{\mathcal{E}'_k^{A_C \rightarrow \tilde{C}}\}_k$  to the  $A_B$  and  $A_C$  parts of the state  $\sigma^{A_B A_C B^n C^n} = (\mathcal{N}(\psi))^{\otimes n}$ . Conditioned on receiving the classical message  $M = m$ , Bob performs  $\mathcal{D}_m^{B^n \rightarrow \tilde{B}}$ . Conditioned on receiving the classical message  $K = k$ , Charlie performs  $\mathcal{D}'_k^{C^n \rightarrow \tilde{C}}$ . By entanglement distillation, there exist  $m$  and  $k$  such that applying  $\mathcal{D}_m \otimes \mathcal{D}'_k$  to  $(\mathcal{E}_m \otimes \mathcal{E}'_k)(\sigma) / \text{Tr}(\mathcal{E}_m \otimes \mathcal{E}'_k)(\sigma)$  gives a state close to the tensor product of the two desired maximally entangled states. Thus, Alice could have prepared

$$\Upsilon_{mk}^{\tilde{B}\tilde{C}A'} = (\mathcal{E}_m \otimes \mathcal{E}'_k)(\psi^{\otimes n}) / \text{Tr}(\mathcal{E}_m \otimes \mathcal{E}'_k)(\psi^{\otimes n})$$

in the first place, eliminating the need for classical communication. If we are interested in entanglement *transmission* instead of entanglement *generation*, then Alice is given a purification of the  $\tilde{B}\tilde{C}$  systems rather than being able to prepare them directly. Luckily, she may always produce  $\Upsilon_{mk}^{\tilde{B}\tilde{C}A'}$  by a (possibly noisy) encoding  $\mathcal{F}$ . A direct adaptation of the result of [1] guarantees that  $\mathcal{F}$  may be replaced by an isometry. It would be desirable to have a direct proof of this theorem (cf. [8] for the single user case) instead of invoking entanglement distillation and [1]. ■

*Remark 3:* The regularized optimization over such  $|\psi\rangle^{ABAC'A'}$  yields the capacity region when there is no Bob-Charlie communication, although the resulting characterization of this capacity region is unlikely to be the most useful.

### III. PROOFS OF MAIN RESULTS

Let us first state some auxilliary results on which our proofs rely.

*Lemma 1 (Gentle measurement (average version) [30]):* Let  $\rho, \Lambda$  be random  $d \times d$  matrices such that  $\rho$  is a density matrix and  $0 \leq \Lambda \leq 1$  which satisfy  $\mathbb{E} \text{Tr} \Lambda \rho \geq 1 - \epsilon$ . Then

$$\mathbb{E} |\sqrt{\Lambda} \rho \sqrt{\Lambda} - \rho|_1 \leq \sqrt{8\epsilon}.$$

*Lemma 2 (Gentle coherent measurement [17]):* Suppose that a POVM  $\{\Lambda_m\}$  identifies the elements of a set of pure states  $\{|\varphi_m\rangle^B\}$ , in the sense that  $\text{Tr} \Lambda_m \varphi_m \geq 1 - \epsilon$  for every  $m$ . Then, there is an isometry  $\mathcal{V}^{B \rightarrow B\tilde{B}}$  which satisfies  $\langle m |^{\tilde{B}} \langle \varphi_m | \mathcal{V} | \varphi_m \rangle \geq 1 - \epsilon$  for each  $m$ .

*Lemma 3 (Continuity Lemma):* If  $\rho^{AB}$  and  $\sigma^{AB}$  satisfy  $|\rho - \sigma|_1 \leq \delta$  for some  $0 \leq \delta \leq 1/e$ , then the following inequalities hold:

$$\begin{aligned} |H(A|B)_\rho - H(A|B)_\sigma| &\leq 2H(\delta) + 4\delta \log |AB| \\ |I(A; B)_\rho - I(A; B)_\sigma| &\leq 3H(\delta) + 6\delta \log |AB|. \end{aligned}$$

*Proof:* Fannes' [13] has shown that

$$|H(\rho^{AB}) - H(\sigma^{AB})| \leq H(\delta) + 2\delta \log |AB|.$$

By monotonicity, the trace distances between partial traces of  $\rho^{AB}$  are no greater than  $\delta$ , so after expanding the conditional entropies and mutual informations, we may apply the triangle inequality, proving the result. ■

We will also need the following two lemmas, which are respectively proved as Lemmas 1 and 2 of [32]:

*Lemma 4:* Suppose  $\rho, \sigma, \Lambda \in \mathcal{B}(\mathcal{H})$ , where  $\rho$  and  $\sigma$  are density matrices, and  $0 \leq \Lambda \leq 1$ . Then,

$$\text{Tr} \Lambda \sigma \geq \text{Tr} \Lambda \rho - |\rho - \sigma|_1.$$

*Lemma 5:* For any state  $\rho^{AB}$  with partial traces  $\rho^A$  and  $\rho^B$  and any  $|\psi\rangle^A$  and  $\sigma^B$ , we have

$$F(\rho^{AB}, \psi^A \otimes \sigma^B) \geq 1 - 3(1 - F(|\psi\rangle^A, \rho^A)) - |\rho^B - \sigma^B|_1.$$

Next, we state an average error version of the HSW Theorem for cq codes with codewords chosen i.i.d. according to a product distribution [27], [20].

*Proposition 2 (HSW Random Coding Theorem):* Given is a cq state  $\sigma^{XQ} = \bigoplus_x p(x) \rho_x^Q$  and a number  $0 \leq R < I(X; Q)_\sigma$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that if  $2^{nR}$  codewords  $\mathcal{C} = \{X^n(m)\}$  are chosen i.i.d. according to the product distribution  $p(x^n) = \prod_{i=1}^n p(x_i)$ , corresponding to input preparations  $\rho_{x^n} = \bigotimes_i \rho_{x_i}$ , there exists a decoding POVM  $\{\Lambda_m\}$  on  $Q^n$ , depending on the random choice of codebook  $\mathcal{C}$ , which correctly identifies the index  $m$  with average probability of error less than  $\epsilon$ , in the sense that

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \rho_{X^n(m)} \Lambda_m \geq 1 - \epsilon. \quad (14)$$

The following proposition is a classical-quantum analog of Corollary 3.8 from [6].

*Proposition 3:* Let  $\{\rho_x^{BC}\}_{x \in \mathcal{X}}$  be a cq channel  $W^{\mathcal{X} \rightarrow BC}$ , and let  $p(x)$  and  $\epsilon, \delta > 0$  be given. If

$$0 \leq R = \min\{I(X; B), I(X; C)\} - \delta$$

and  $n$  is large enough, there is a set of  $2^{nR}$  HSW codewords  $\{x^n(m)\}$ , each of the same type  $P$  satisfying  $|P - p|_1 \leq \delta$ , a measurement on  $B^n$  with POVM  $\{\Lambda_m\}$  and a measurement on  $C^n$  with POVM  $\{\Lambda'_m\}$  such that for each  $m$ ,

$$\text{Tr}(\Lambda_m \otimes \Lambda'_m) \rho_m \geq 1 - \epsilon$$

where  $\rho_m = \bigotimes_i \rho_{x_i(m)}$ .

*Proof:* Follows from standard arguments (see e.g. [29]). ■

The following quantum coding proposition for single-user channels is proved in [8] and concerns the existence of random entanglement transmission codes whose average code density matrix can be made arbitrarily close to a product state.

*Proposition 4 (LSD Random Coding Theorem):* Given is a channel  $\mathcal{N}: A' \rightarrow B$ , a density matrix  $\rho^{A'}$ , and a number  $0 \leq R < I_c(\rho, \mathcal{N})$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that there is a random ensemble of  $(2^{nR}, n, \epsilon)$  entanglement generation codes  $(p_\beta, |\Upsilon_\beta\rangle^{AA'n}, \mathcal{D}_\beta^{B^n \rightarrow \hat{A}})$  for  $\mathcal{N}$  with average code density operator  $\varrho^{A'n} = \sum_\beta p_\beta \text{Tr}_A \Upsilon_\beta$  satisfying  $|\varrho - \rho^{\otimes n}|_1 \leq \epsilon$ . Moreover, each code in the ensemble is good, in the sense that  $F(|\Phi\rangle^{A\hat{A}}, \mathcal{D}_\beta \circ \mathcal{N}^{\otimes n}(\Upsilon_\beta^{AA'n})) \geq 1 - \epsilon$  for each value of the randomness  $\beta$ .

*Remark 4:* In practice, we omit the common randomness index, treating the encoding and decoding as a pair of correlated random objects.

#### A. Proof of Theorem 1

*Proof: (Coding theorem)* Let  $W^{\mathcal{X} \rightarrow BC}$  be a cq broadcast channel with conditional density matrices  $\rho_x^{BC}$  and let  $p(t, x)$  be arbitrary. Together, these probabilities and states define the joint cq state

$$\sigma^{TXBC} = \bigoplus_{t,x} p(t, x) \rho_x^{BC} \equiv \bigoplus_t p(t) \sigma_t^{XBC}.$$

The corresponding conditional distribution  $p(x|t)$  defines a set of conditional density matrices

$$\tau_t^{BC} = \sum_x p(x|t) \rho_x^{BC} = \text{Tr}_X \sigma_t^{XBC}$$

for a new cq channel  $V^{T \rightarrow BC}$ , representing a “backed up” version of the original channel  $W$ . Note that these conditional density matrices can be used to rewrite

$$\sigma^{TBC} = \text{Tr}_X \sigma^{TXBC} = \bigoplus_t p(t) \tau_t^{BC}.$$

For any  $\epsilon, \delta > 0$  and sufficiently large  $n$ , we will show that for rates  $R_B$  and  $R$  satisfying

$$I(X; B|T)_\sigma - (1 + |\mathcal{X}|)\delta \leq R_B < I(X; B|T)_\sigma$$

and

$$0 \leq R = \min\{I(T; B)_\sigma, I(T; C)_\sigma\} - \delta,$$

there exists an  $(R, R_B, n, 2\epsilon^{1/8})$  code for  $W^{\mathcal{X} \rightarrow BC}$ .

We will construct the required doubly-indexed set of codewords  $\{x(m, k)\}_{m \in 2^{nR}, k \in 2^{nR_B}}$  as follows. First, we select a rate  $R$  code for the channel  $V^{T \rightarrow BC}$  which conveys the index  $m \in 2^{nR}$  to Bob and Charlie. Then, for each  $t$ , we pick a random HSW code of blocklength approximately  $p(t)n$  for  $W^{\mathcal{X} \rightarrow BC}$  with codewords selected i.i.d. according to  $p(x|t)$ , such that if Bob knows  $t$ , he can decode at rates approaching  $I(X; B)_\sigma$ . Note that because of the randomness in this second coding layer, the average state seen by Bob on any channel output where the  $t$ 'th code was used is equal to  $\tau_t^{BC}$ .

To decode, Bob and Charlie first use their measurements from the common code, allowing them to identify  $m$  well on average. In addition to knowing the common message  $m$ , Bob then knows which instances of the channel were used with which random codes, so that he can apply an appropriate decoder, which depends on the randomness in the second coding layer, to learn his personalized message  $k$ . Note that since

$$I(X; B|T)_\sigma = \sum_t p(t) I(X; B)_\sigma,$$

the personal rate to Bob will be near that which is desired. We then infer the existence of a deterministic code with low error probability for all message pairs.

We begin by invoking Proposition 3 to obtain an  $(R, n, \epsilon)$  code  $\{t^n(m), \Lambda_m, \Lambda'_m\}_{m \in 2^{nR}}$  for  $V^{T \rightarrow BC}$  with codewords of type  $P$  satisfying  $|P - p|_1 \leq \delta$ . Recall that for each  $m$ ,

$$\text{Tr}(\Lambda_m \otimes \Lambda'_m) \tau_m^{B^n C^n} \geq 1 - \epsilon, \quad (15)$$

where  $\tau_m = \bigotimes_i \tau_{t_i(m)}$ .

For each  $t$ , define the integer  $n_t = nP(t)$ , as well as  $\epsilon_t = \epsilon P(t)$ ,  $\delta_t = \delta P(t)$  and  $R_t = I(X; B)_\sigma - \delta_t \leq |\mathcal{X}|$ . It follows from Proposition 2 that for each  $t$ , there exists an  $(R_t, n_t, \epsilon_t)$  random HSW code  $\{X^{n_t}(k_t|t), \Lambda_k^{(t)}\}_{k_t \in 2^{nR_t}}$  (here,  $\{X^{n_t}(k_t|t)\}$  is just a doubly indexed family of random variables) for the channel  $W^{\mathcal{X} \rightarrow B}$  to Bob which satisfies

$$\mathbb{E} 2^{-nR_t} \sum_{k_t=1}^{2^{nR_t}} \text{Tr} \rho_{k_t}^{B_t} \Lambda_k^{(t)} \geq 1 - \epsilon_t \quad (16)$$

where the expectation is over the randomness in the HSW codes. Above, we have abbreviated  $B_t \equiv B^{n_t}$  and taken

$$\rho_{k_t}^{B_t} = \bigotimes_{i=1}^{n_t} \rho_{X_i(k_t|t)}^{B_i}.$$



Each  $X_i(k_t|t)$  is chosen independently according to  $p(x|t)$ , so that  $\mathbb{E} \rho_{k_t} = \tau_t^{\otimes n_t}$ . Observe that by the symmetry of the random code construction, (16) may be equivalently expressed as

$$\mathbb{E} \text{Tr} \rho_1^{B_t} \Lambda_1^{(t)} \geq 1 - \epsilon_t.$$

Noting that the personal rate to Bob is given by

$$\begin{aligned} R_B &= \sum_t \frac{n_t}{n} R_t = \sum_t P(t) R_t \\ &\geq \sum_t p(t) R_t - |P - p|_1 |\mathcal{X}| \\ &\geq I(X; B|T) - (|\mathcal{X}| + 1)\delta, \end{aligned}$$

and also that  $R_B < I(X; B|T)$ , we may uniquely identify any message  $k \in 2^{nR_B}$  for Bob with a collection of messages  $\{k_t \in 2^{nR_t}\}_t$ . Recalling that all of the codewords  $\{t^n(m)\}_{m \in 2^{nR}}$  are of the same type and setting  $d = |T|$ , we may assume w.l.o.g. that  $t^n(1) = 1^{n_1} 2^{n_2} \dots d^{n_d}$ , so that we may identify a collection of permutations  $\{\pi(m): T^n \rightarrow T^n\}$  for which  $t^n(m) = \pi(m)(t^n(1))$ . By letting these permutations act on  $\mathcal{X}^n$  in the same way, we may define Alice's (random) encoding via

$$X^n(m, k) = \pi(m)(X^{n_1}(k_1|1)X^{n_2}(k_2|2) \dots X^{n_d}(k_d|d)).$$

We abbreviate  $\rho_{mk}^{B^n C^n} = \rho_{X^n(mk)}^{B^n C^n}$ , observing that for each  $k$ , we have  $\mathbb{E} \rho_{mk} = \tau_m$ .

To decode, Bob first measures  $\{\Lambda_m\}$  while Charlie measures  $\{\Lambda'_m\}$ , after which they declare their respective results to be the common message  $M$ . Next, Bob will permute his  $B^n$  systems according to  $\pi^{-1}(m)$ , obtaining a state close to  $\rho_{1m}^{B^n}$ . For each  $t$ , he then measures each block of  $n_t$  outputs with the corresponding  $\{\Lambda_{k_t}^{(t)}\}$  to obtain  $(k_1, \dots, k_t) = k$ , which he declares as his personal message. Bob's overall procedure can be summarized in terms of the POVM  $\{\Lambda_{mk}\}$ , defined as  $\Lambda_{mk} = \sqrt{\Lambda_m} \Lambda_{k|m} \sqrt{\Lambda_m}$ , where we take

$$\Lambda_{k|m} = \pi(m) \left( \bigotimes_t \Lambda_{k_t}^{(t)} \right)$$

with  $\pi(m)$  now acting to permute  $B^n$  in the obvious way. Defining

$$\begin{aligned} P_{mk} &= \text{Tr}(\Lambda_{mk} \otimes \Lambda'_m) \rho_{mk}^{B^n C^n}, \\ \tilde{\rho}_{mk} &= (\sqrt{\Lambda_m} \otimes \sqrt{\Lambda'_m}) \rho_{mk}^{B^n C^n} (\sqrt{\Lambda_m} \otimes \sqrt{\Lambda'_m}), \end{aligned}$$

we estimate

$$\begin{aligned} \mathbb{E} P_{mk} &= \mathbb{E} \text{Tr} \tilde{\rho}_{mk}^{B^n C^n} \Lambda_{k|m} \\ &\geq \mathbb{E} \text{Tr} \rho_{mk}^{B^n} \Lambda_{k|m} - \mathbb{E} |\tilde{\rho}_{mk}^{B^n C^n} - \rho_{mk}^{B^n C^n}|_1 \\ &\geq \mathbb{E} \text{Tr} \rho_{mk}^{B^n} \Lambda_{k|m} - \sqrt{8\epsilon} \\ &= \mathbb{E} \text{Tr} \left( \bigotimes_t \rho_1^{B_t} \right) \left( \bigotimes_t \Lambda_1^{(t)} \right) - \sqrt{8\epsilon} \\ &= \prod_t \mathbb{E} \text{Tr} \rho_1^{B_t} \Lambda_1^{(t)} - \sqrt{8\epsilon} \\ &\geq 1 - \sum_t \epsilon_t - \sqrt{8\epsilon} \\ &\geq 1 - 4\sqrt{\epsilon}. \end{aligned}$$

The first inequality is by Lemma 4 and the second by Lemma 1. We may now derandomize, concluding that there is a particular value of the common randomness such that

$$\begin{aligned} 2^{-n(R_B+R)} \sum_{m=1}^{2^{nR}} \sum_{k=1}^{2^{nR_B}} P_{mk} &\equiv 2^{-nR} \sum_{m=1}^{2^{nR}} \bar{P}_m \\ &\geq 1 - 4\sqrt{\epsilon} \end{aligned}$$

According to Markov's inequality, half of the values of  $m$  satisfy  $\bar{P}_m \geq 1 - 2\epsilon^{1/4}$ . Among those, half of the corresponding  $k$ 's are such that  $P_{mk} \geq 1 - \sqrt{1 - \bar{P}_m} \geq 1 - 2\epsilon^{1/8}$ . By only using those  $m$ 's, the common rate  $R$  is reduced by a negligible  $\frac{1}{n}$ . For each such  $m$ , throwing out the worst half of the  $k$ 's reduces  $R_B$  by the same amount. This completes the proof. ■

*Proof: (Converse)* Assuming that  $(R_B, R)$  is achievable, let  $\{x^n(m, k)\}$ ,  $\{\Lambda_{mk}\}$  and  $\{\Lambda'_m\}$  comprise any  $(R, n, \epsilon_n)$  code in the achieving sequence. Setting

$$\Pi_{mk}^{MKX^n} = |m\rangle\langle m| \otimes |k\rangle\langle k| \otimes |x^n(m, k)\rangle\langle x^n(m, k)|,$$

we write

$$\omega^{MKX^n B^n C^n} = 2^{-n(R_B+R)} \sum_{m=1}^{2^{nR}} \sum_{k=1}^{2^{nR_B}} \Pi_{mk} \otimes \rho_{mk}^{B^n C^n}.$$

Supposing that Bob stores his decoded messages in the registers  $K_B$  and  $M_B$ , while Charlie stores his in  $M_C$ , let  $\Omega^{MKM_B M_C K_B}$  be the joint state after the decoding. Then, for some  $\epsilon'_n, \epsilon''_n, \epsilon'''_n \rightarrow 0$ , we have

$$\begin{aligned} nR &= H(M) \\ &\leq I(M; M_C)_\Omega + n\epsilon'_n \\ &\leq I(M; C^n)_\omega + n\epsilon'_n, \end{aligned} \quad (17)$$

The second line is by Fano's inequality (see e.g. [5]) and the third is by the Holevo bound [19]. A similar argument yields  $nR \leq I(M; B^n) + n\epsilon''_n$ . Finally, we bound

$$\begin{aligned} nR_B &= H(K) \\ &\leq I(K; K_B)_\Omega + n\epsilon'''_n \\ &\leq I(K; B^n)_\omega + n\epsilon'''_n \\ &\leq I(K; B^n M)_\omega + n\epsilon'''_n \\ &= I(K; B^n | M)_\omega + n\epsilon'''_n \\ &\leq I(X^n; B^n | M)_\omega + n\epsilon'''_n. \end{aligned} \quad (18)$$

The middle four lines are by Fano's inequality, the Holevo bound, data processing, and the independence of  $K$  and  $M$ . The last inequality uses the Markov chain  $KM - X^n - B^n$ . Choosing  $T \equiv M$  completes the proof. ■

## B. Proof of Theorem 2

By data processing,  $I(X; C)_\sigma \leq I(X; B)_\sigma$ . Hence, the coding theorem follows from the previous one. It only remains to single-letterize the bounds appearing in the previous converse.



*Proof: (Converse)* We begin by rewriting the conditional information from (18):

$$\begin{aligned}
I(X^n; B^n | M) &= H(B^n | M) - H(B^n | X^n M) \\
&= \sum_{i=1}^n [H(B_i | B^{i-1} M) - H(B_i | X^n B^{i-1} M)] \\
&= \sum_{i=1}^n [H(B_i | B^{i-1} M) - H(B_i | X_i B^{i-1} M)] \\
&= \sum_{i=1}^n I(X_i; B_i | M B^{i-1}) \\
&= nI(X_S; B_S | M B^{S-1} S) \\
&= nI(X; B | T).
\end{aligned}$$

The third line holds because of the Markov chain

$$X^{i-1} X_{i+1}^n - X_i M B^{i-1} - B_i,$$

where we abbreviate  $X_{i+1}^n = X_{i+1} \cdots X_n$  for  $i < n$ , setting it equal to a constant when  $i = n$ . To see that this is a Markov chain, note that the left recovery map is deterministic, while the right recovery map prepares the appropriate state of  $B_i$  given the value of  $X_i$ . In the remaining steps, we define  $S \sim \text{unif}\{1, \dots, n\}$ . The last identifies  $T \equiv S M B^{S-1}$  and  $X \equiv X_S$ . We continue bounding (17):

$$\begin{aligned}
I(M; C^n) &= \sum_{i=1}^n I(M; C_i | C^{i-1}) \\
&= \sum_{i=1}^n [H(C_i | C^{i-1}) - H(C_i | M C^{i-1})] \\
&\leq \sum_{i=1}^n [H(C_i) - H(C_i | M C^{i-1})] \\
&\leq \sum_{i=1}^n [H(C_i) - H(C_i | M B^{i-1})] \\
&= \sum_{i=1}^n I(M B^{i-1}; C_i) \\
&= nI(M B^{S-1}; C_S | S) \\
&\leq n[I(M B^{S-1}; C_S | S) + I(S; C_S)] \\
&= nI(S M B^{S-1}; C_S) \\
&= nI(T; C).
\end{aligned} \tag{19}$$

Here, the third and fourth lines follow from the fact that conditioning reduces entropy and data processing with respect to appropriate tensor products of the degrading map  $\mathcal{M}^{B \rightarrow C}$ . The last step identifies  $C \equiv C_S$ . Observe that the commutativity of the  $\{\rho_x^B\}$  was needed to identify  $T$  with a classical random variable. ■

### C. Proof of Theorem 3

*Proof: (Coding theorem)* Let  $\mathcal{N}^{A' \rightarrow BC}$  be an arbitrary broadcast channel and fix an ensemble of bipartite pure states  $\{p(t), |\phi_t\rangle^{A'' A'}\}$ . For any  $\epsilon, \delta > 0$  and sufficiently large  $n$ , we will show that there exists an  $(R, Q, n, 3\epsilon^{1/4})$  cq entanglement generation code

$$\{|\Upsilon_m\rangle^{AA'}, \mathcal{D}_1^{B^n \rightarrow M_B \hat{A}}, \mathcal{D}_2^{C^n \rightarrow M_C}\}_{m \in 2^{nR}}$$

for  $\mathcal{N}^{A' \rightarrow BC}$ , provided that  $0 \leq Q = I(A'' \rangle BT)_\sigma - \delta$  and  $0 \leq R = \min\{I(T; B)_\sigma, I(T; C)_\sigma\} - \delta$ , where

$$\sigma^{TABC} = \bigoplus_t p(t) \mathcal{N}(\phi_t^{A'' A'}).$$

We do this by showing that there are two POVMs:  $\{\Lambda_m\}_{m \in 2^{nR}}$  on  $B^n$  and  $\{\Lambda'_m\}_{m \in 2^{nR}}$  on  $C^n$ , as well as a collection of maps  $\mathcal{D}_m^{B^n \rightarrow \hat{A}}$ , for which the trace-reducing maps  $\{\mathcal{D}_m(\sqrt{\Lambda_m}(\cdot)\sqrt{\Lambda_m})\}_{m \in 2^{nR}}$  are the components of  $\mathcal{D}_1$ , and  $\mathcal{D}_2$  implements  $\{\Lambda'_m\}_{m \in 2^{nR}}$ .

For each  $t$ , we set  $\rho_t^{A'} = \text{Tr}_{A''} \phi_t$  and  $\tau_t^{BC} = \mathcal{N}(\rho_t)$ , defining a cq channel  $V^{T \rightarrow BC}$  with conditional density matrices  $\tau_t^{BC}$ . As in the previous coding theorem, we invoke Proposition 3 to obtain, for sufficiently large  $n$ , an  $(R, n, \epsilon)$  code  $\{t^n(m), \Lambda_m, \Lambda'_m\}_{m \in 2^{nR}}$  for  $V$  with codewords of type  $P$  satisfying  $|P - p|_1 \leq \epsilon$ . For each  $m$ , we write  $\rho_m^{A'n} \equiv \bigotimes_i \rho_{t_i(m)}^{A'}$ , recalling that

$$\text{Tr}(\Lambda_m \otimes \Lambda'_m) \mathcal{N}^{\otimes n}(\rho_m^{A'n}) \geq 1 - \epsilon. \tag{20}$$

As in the direct coding part of the proof of Theorem 1 we define  $n_t = nP(t)$ ,  $\epsilon_t = \epsilon P(t)$  and  $\delta_t = \delta P(t)$ . We also assume that for  $|T| = d$ , the first codeword is  $t^n(1) = 1^{n_1} 2^{n_2} \cdots d^{n_d}$  so that there are permutations  $\pi(m)$  of  $\mathcal{T}^n$  satisfying  $t^n(m) = \pi(m)(t^n(1))$ .

For each  $t \in \mathcal{T}$ , we may set  $Q_t = I_c(\tau_t, \mathcal{N}) - \delta_t$  and conclude from Proposition 4 that there exists a  $(Q_t, n_t, \epsilon_t)$  random entanglement generation code  $\{|\Upsilon_t\rangle^{A_t A'^{n_t}}, \mathcal{D}_t^{B^{n_t} \rightarrow \hat{A}_t}\}$  whose average code density operator  $\varrho_t^{A'^{n_t}} = \mathbb{E} \text{Tr}_{A_t} \Upsilon_t$  satisfies

$$|\varrho_t^{A'^{n_t}} - \rho_t^{\otimes n_t}|_1 \leq \epsilon_t. \tag{21}$$

It is also guaranteed that for each  $t$ , the state

$$\xi_t^{A_t B^{n_t} C^{n_t}} \equiv \mathcal{N}^{\otimes n}(\Upsilon_t)$$

created by the  $t$ th random quantum code approximately contains rate  $Q_t$  entanglement between Alice and Bob, in the sense that

$$F(|\Phi_{Q_t}\rangle^{A_t \hat{A}_t}, \text{Tr}_{C^{n_t}} \mathcal{D}_t(\xi_t)) \geq 1 - \epsilon_t. \tag{22}$$

Equating  $A \equiv \bigotimes_t A_t$ , we make the definitions

$$\begin{aligned}
|\Upsilon_1\rangle^{AA'} &= \bigotimes_t |\Upsilon_t\rangle^{A_t A'^{n_t}} \\
|\Upsilon_m\rangle^{AA'} &= (1^A \otimes \pi(m)) |\Upsilon_1\rangle^{AA'},
\end{aligned}$$

where we extend  $\pi(m)$  to act by permuting the registers  $A'^n$  in the obvious way. Defining the average code density operator for the new code as  $\varrho_m^{A'n} = \mathbb{E} \text{Tr}_A \Upsilon_m$ , note that we can bound

$$\begin{aligned}
|\varrho_m - \rho_m|_1 &= \left| \bigotimes_t \varrho_t^{A'^{n_t}} - \bigotimes_t \rho_t^{\otimes n_t} \right|_1 \\
&\leq \sum_t |\varrho_t^{A'^{n_t}} - \rho_t^{\otimes n_t}|_1 \\
&\leq \sum_t \epsilon_t \\
&= \epsilon,
\end{aligned} \tag{23}$$

where we have used unitary invariance of the trace norm, telescoping, and (21), in that order. To send the classical

message  $m$ , Alice prepares the state  $|\Upsilon_m\rangle^{AA'^n}$ . The structure of the decoder is similar to that from the proof of Theorem 1. Bob and Charlie begin by performing their respective measurements, in order to ascertain the classical message. Then Bob permutes his output systems accordingly and applies the quantum decoder  $\mathcal{D} \equiv \bigotimes_t \mathcal{D}_t$ .

We will write the the joint state after Alice sends her encoding through the channel as

$$\vartheta_m^{AB^n C^n} = \mathcal{N}^{\otimes n}(\Upsilon_m^{AA'^n}),$$

so that in particular, the state corresponding to the first message is  $\vartheta_1^{AB^n C^n} = \bigotimes_t \xi_t^{A_t B^{n_t} C^{n_t}}$ . Note that if the decoder  $\mathcal{D}^{B^n \rightarrow \hat{A}}$  is applied directly to  $\vartheta_1$ , the resulting Alice-Bob state is nearly maximally entangled:

$$\begin{aligned} F(|\Phi_Q\rangle^{A\hat{A}}, \text{Tr}_{C^n} \mathcal{D}(\vartheta_1)) &= \prod_t F(|\Phi_{Q_t}\rangle^{A_t \hat{A}_t}, \text{Tr}_{C^{n_t}} \mathcal{D}_t(\xi_t)) \\ &\geq \prod_t (1 - \epsilon_t) \\ &\geq 1 - \sum_t \epsilon_t \\ &= 1 - \epsilon, \end{aligned} \quad (24)$$

where the first inequality is by (22). Next, define the subnormalized density matrices

$$\tilde{\vartheta}_m^{AB^n C^n} = (\sqrt{\Lambda_m} \otimes \sqrt{\Lambda'_m}) \vartheta_m^{AB^n C^n} (\sqrt{\Lambda_m} \otimes \sqrt{\Lambda'_m})$$

which depend on the shared randomness in the quantum code, and are proportional to the states which result when Bob and Charlie both correctly learn the message  $m$ . This happens with probability bounded as

$$\begin{aligned} \mathbb{E} \text{Tr} \tilde{\vartheta}_m^{AB^n C^n} &= \text{Tr}(\Lambda_m \otimes \Lambda'_m) \mathbb{E} \text{Tr}_A \vartheta_m^{AB^n C^n} \\ &= \text{Tr}(\Lambda_m \otimes \Lambda'_m) \mathcal{N}^{\otimes n}(\varrho^{A'^n}) \\ &\geq \text{Tr}(\Lambda_m \otimes \Lambda'_m) \mathcal{N}^{\otimes n}(\rho^{A'^n}) \\ &\quad - |\varrho^{A'^n} - \rho^{A'^n}|_1 \\ &\geq 1 - 2\epsilon. \end{aligned} \quad (25)$$

In the second to last line, we have applied Lemma 4 along with monotonicity with respect to  $\mathcal{N}^{\otimes n}$ , while the last line uses the estimates (20) and (23). We may now write the expectation, over the shared randomness in the quantum code, of the fidelity  $F_m$  between the state resulting from the protocol when the  $m$ th common message is sent and the target maximally entangled state as

$$\begin{aligned} \mathbb{E} F_m &= \mathbb{E} F_1 \\ &\equiv \mathbb{E} F(|\Phi_Q\rangle^{A\hat{A}}, \text{Tr}_{C^n} \mathcal{D}(\tilde{\vartheta}_1^{AB^n C^n})) \\ &\geq 1 - \mathbb{E} |\Phi_Q - \text{Tr}_{C^n} \mathcal{D}(\tilde{\vartheta}_1)|_1 \\ &\geq 1 - \mathbb{E} |\Phi_Q - \text{Tr}_{C^n} \mathcal{D}(\vartheta_1)|_1 - \mathbb{E} |\vartheta_1 - \tilde{\vartheta}_1|_1 \\ &\geq 1 - 2\sqrt{\epsilon} - \sqrt{8 \cdot 2\epsilon} \\ &\geq 1 - 6\sqrt{\epsilon}. \end{aligned}$$

Here, the first line follows by the permutation symmetry of the code, while the third uses (3). The fourth is a consequence of the triangle inequality, together with monotonicity with respect

to  $\text{Tr}_{C^n} \mathcal{D}$ . The estimates in the second to last line are obtained by applying (4) to (24) (which holds without the expectation), as well as Lemma 1 to (25).

At this point, it is possible to derandomize our code. Having proved that

$$\mathbb{E} 2^{-nR} \sum_m F_m \geq 1 - 6\sqrt{\epsilon},$$

we may conclude that there is a deterministic value of the shared randomness from the quantum codes yielding the same average error bound. By throwing out the worst half of the codewords, Markov's inequality implies that we are left with a code for which

$$F_m \geq 1 - 3\epsilon^{1/4}$$

for every  $m$ , while reducing the rate by a negligible  $\frac{1}{n}$ . ■

*Proof: (Converse)* Assume that  $(Q_B, R)$  is achievable and let  $\{|\Upsilon_m\rangle^{AA'^n}\}_{m \in 2^{nR}}$ ,  $\mathcal{D}_1^{B^n \rightarrow \hat{A} M_B}$  and  $\mathcal{D}_2^{C^n \rightarrow M_C}$  be a  $(Q, R, n, \epsilon_n)$  cq entanglement generation code from any achieving sequence. Defining the state

$$\omega^{MAB^n C^n} = 2^{-nR} \bigoplus_{m \in 2^{nR}} \mathcal{N}^{\otimes n}(\Upsilon_m^{AA'^n}) \quad (26)$$

and setting  $\Omega^{MM_B M_C A\hat{A}} = (\mathcal{D}_1 \otimes \mathcal{D}_2)(\omega)$ , we may upper bound the quantum rate  $Q$  via

$$\begin{aligned} I(A)B^n M)_\omega &\geq I(A)\hat{A})_\Omega \\ &\geq I(A)\hat{A})_{\Phi_{Q_B}} - n\epsilon'_n \\ &= nQ_B - n\epsilon'_n. \end{aligned} \quad (27)$$

The first step is by data processing with respect to  $\text{Tr}_M \mathcal{D}_1$ , while the second is by the Continuity Lemma 3, for some  $\epsilon'_n \rightarrow 0$ . The classical rate  $R$  may also be bounded as

$$\begin{aligned} nR &= H(M)_\Omega \\ &\leq I(M; M_C)_\Omega + n\epsilon''_n \\ &\leq I(M; C^m)_\omega + n\epsilon''_n, \end{aligned} \quad (28)$$

where  $\epsilon''_n \rightarrow 0$ , and we have used Fano's inequality and the Holevo bound. Another consequence of the Holevo bound is that  $I(M; M_B)_\Omega \leq I(M; B^n)_\omega$ , yielding  $nR \leq \min\{I(M; B^n)_\omega, I(M; C^m)_\omega\}$ . We have thus shown that for any  $\delta > 0$ , the rate pair  $(Q - \delta, R - \delta)$  is contained in  $\mathcal{Q}(\mathcal{N})$ . As  $\mathcal{Q}(\mathcal{N})$  is closed by definition, this completes the proof. ■

#### D. Proof of Theorem 4

As the coding theorem follows from that of Theorem 3, we need only single-letterize the corresponding multi-letter converse.

*Proof: (Converse)* Under the assumption that  $\mathcal{N}^{A' \rightarrow B}$  is a generalized dephasing channel, we will further upper bound the information quantities (27) and (28) appearing in the multi-letter converse of Section III-C by appropriate single-letter quantities. We begin working with the state  $\omega^{MAB^n C^n}$  from (26) which is induced by an  $(R, Q, n, \epsilon_n)$  cq entanglement generation code from an achieving sequence. Recalling from (5) that the completely dephasing channel  $\Delta$  sets to zero all

off-diagonal matrix elements in the dephasing basis  $\{|x\rangle\}$ , set  $\varrho_m^{A'n} = \Delta^{\otimes n}(\text{Tr}_A \Upsilon_m)$ , observing that we may write

$$\varrho_m^{A'n} = \bigoplus_{x^n} p(x^n|m)$$

for some conditional probabilities  $p(x^n|m)$ . Let us now define the state

$$\begin{aligned} \omega'^{MB^n C^n E^n} &= 2^{-nR} \bigoplus_{m \in 2^{nR}} \mathcal{U}^{\otimes n}(\varrho_m) \\ &= 2^{-nR} \bigoplus_{m \in 2^{nR}} \sum_{x^n} p(x^n|m) \psi_{x^n}^{C^n E^n} \end{aligned}$$

where we abbreviate  $\psi_{x^n}^{C^n E^n} \equiv \bigotimes_i \psi_{x_i}^{C_i E_i}$ . Abbreviating  $\mathcal{N}^{A' \rightarrow B}$  to  $\mathcal{N}_B$  and  $\mathcal{N}^{A' \rightarrow C}$  to  $\mathcal{N}_C$ , the left hand side of (27) can be written

$$I(A)B^n M)_\omega = 2^{-nR} \sum_{m \in 2^{nR}} I_c(\text{Tr}_A \Upsilon_m, \mathcal{N}_B^{\otimes n}).$$

By (6) and (7), each summand can be upper bounded as

$$I_c(\text{Tr}_A \Upsilon_m, \mathcal{N}_B^{\otimes n}) \leq H(\mathcal{N}_B^{\otimes n}(\varrho_m)) - H((\mathcal{N}_B)_c^{\otimes n}(\varrho_m)).$$

Combining these last two equations yields

$$\begin{aligned} I(A)B^n M)_\omega &\leq H(B^n|M)_{\omega'} - H(C^n E^n|M)_{\omega'} \\ &= H(X^n|M)_{\omega'} - H(C^n E^n|M)_{\omega'} \end{aligned}$$

where we have renamed  $B^n$  to  $X^n$  to emphasize its classicality. From now on, we rename  $\omega'^{MB^n C^n E^n}$  to  $\omega'^{MX^n C^n E^n}$  accordingly. Identifying  $T_i = MX^{i-1}$ ,  $T = ST_S$  and  $XCE = X_S C_S E_S$ , for  $S \sim \text{unif}\{1, \dots, n\}$ , observe that  $S - T - XCE$  forms a Markov chain. This identification defines the state  $\Omega^{TXCE}$ , for which

$$\begin{aligned} H(X^n|M)_{\omega'} &= \sum_{i=1}^n H(X_i|MX^{i-1})_{\omega'} \\ &= nH(X|T)_\Omega. \end{aligned}$$

By data processing with respect to appropriate tensor products of the map  $|x\rangle\langle x| \mapsto \psi_x^{CE}$ , we may upper bound

$$\begin{aligned} -H(C^n E^n|M)_{\omega'} &= -\sum_{i=1}^n H(E_i C_i | E^{i-1} C^{i-1} M)_{\omega'} \\ &\leq -\sum_{i=1}^n H(C_i E_i | MX^{i-1})_{\omega'} \\ &= -nH(CE|T)_\Omega, \end{aligned}$$

obtaining  $\frac{1}{n}I(A)B^n M)_\omega \leq H(X|T)_\Omega - H(CE|T)_\Omega$ .

It is perhaps instructive to see that  $\Omega$  can be explicitly written as  $\Omega^{TXCE} = \bigoplus_t p(t) \Omega_t^{XCE}$ , where we take  $\mathcal{T} = \mathcal{M} \times \biguplus_s \mathcal{X}^{s-1}$  (here  $\mathcal{X}^0$  is the empty set), and

$$\Omega_{mx^{s-1}}^{XCE} = \bigoplus_{x_s} p(x_s|m, x^{s-1}) \psi_{x_s}^{CE}.$$

We now continue by bounding the mutual information in (28) via

$$\begin{aligned} I(M; C^n)_\omega &= I(M; C^n)_{\omega'} \\ &\leq nI(T; C)_\Omega. \end{aligned}$$

Here, the first step is because  $\mathcal{N}_C^{\otimes n} \circ \Delta^{\otimes n} = \mathcal{N}_C^{\otimes n}$ , which follows from (6) because  $\mathcal{N}_C = \text{Tr}_E(\mathcal{N}_B)_c$ , while the second follows from manipulations which are identical to those used to bound (19) in the converse to Theorem 2; the only differences are that we use data processing with respect to tensor products of the map  $|x\rangle\langle x| \mapsto \text{Tr}_E \psi_x$  and relabel  $B^{i-1}$  to  $X^{i-1}$ . This proves the claim. ■

### E. Proof of Theorem 5

*Proof: (Coding theorem)* Letting  $\mathcal{U}^{A' \rightarrow BC}$  be an arbitrary isometry, we set  $\mathcal{N}_B = \text{Tr}_C \mathcal{U}$ . For any bipartite pure state ensemble  $\{p(t), |\phi_t\rangle^{A'' A'^n}\}_{m \in 2^{nR}}$  and any  $\epsilon > 0$ , the previous coding theorem shows (relabeling  $R$  to  $Q$  and  $Q$  to  $Q_B$ ) that as long as  $n$  is large enough, there is a  $(Q, Q_B, n, 6\sqrt{\epsilon})$  cq entanglement generation code  $\{\Upsilon_m\}, \mathcal{D}_m, \{\Lambda_m\}, \{\Lambda'_m\}$  for  $\mathcal{U}^{A' \rightarrow BC}$ , provided that the rates satisfy

$$Q < \min\{I(T; B)_\sigma, I(T; C)_\sigma\}$$

and

$$Q_B < I(A'' \rangle BT)_\sigma.$$

These quantities are computed with respect to the state

$$\sigma^{A'' BCT} = \bigoplus_t p(t) \mathcal{U}(\phi_t^{A'' A'}).$$

We will show how to make the common classical message coherent. For each  $m \in 2^{nR}$ , define

$$|\Upsilon'_m\rangle^{AB^n C^n} = \mathcal{U}^{\otimes n} |\Upsilon_m\rangle$$

and observe that

$$\langle \Upsilon'_m | (1^A \otimes \Lambda_m \otimes \Lambda'_m) | \Upsilon'_m \rangle \geq 1 - \epsilon.$$

By Lemma 2, there are thus coherent local measurements  $\mathcal{V}^{B^n \rightarrow B^n G_B}$  and  $\mathcal{W}^{C^n \rightarrow C^n G_C}$  satisfying

$$\langle m |^{G_C G_B} \langle \Upsilon'_m | (\mathcal{V} \otimes \mathcal{W}) | \Upsilon'_m \rangle \geq 1 - \epsilon \quad (29)$$

for each  $m$ , where we take  $|m\rangle^{G_B G_C} \equiv |m\rangle^{G_B} |m\rangle^{G_C}$ . Now, there are local unitaries (permutations of the Hilbert space factors, in fact)  $V_m^{B^n \rightarrow B^n}$  and  $W_m^{C^n \rightarrow C^n}$  which satisfy

$$(V_m \otimes W_m) | \Upsilon'_m \rangle = | \Upsilon'_1 \rangle^{AB^n C^n}$$

because  $|\Upsilon_m\rangle$  is just a permutation of the  $A'^n$  part of the fixed representative  $|\Upsilon_1\rangle^{AA'^n}$ . Define the controlled unitary

$$V^{B^n G_B \rightarrow B^n G_B} = \sum_m |m\rangle\langle m| \otimes V_m$$

and similarly define  $W^{C^n G_C \rightarrow C^n G_C}$ . Setting

$$|\Upsilon''_m\rangle^{AB^n C^n G_B G_C} = ((V \circ \mathcal{V}) \otimes (W \circ \mathcal{W})) | \Upsilon'_m \rangle,$$

we may reexpress (29) as

$$\langle m |^{G_B G_C} \langle \Upsilon'_1 | | \Upsilon''_m \rangle \geq 1 - \epsilon. \quad (30)$$

We now define Alice's encoding as

$$|\Upsilon\rangle^{GAA'^n} = \frac{1}{\sqrt{2^{nQ}}} \sum_m |m\rangle^G |\Upsilon_m\rangle^{AA'^n},$$

writing

$$|\Upsilon'\rangle^{GAB^n C^n} = \mathcal{U}^{\otimes n} |\Upsilon\rangle^{AA'^n}$$

and also setting

$$|\Upsilon''\rangle^{GAB^n C^n G_B G_C} = ((V \circ \mathcal{V}) \otimes (W \circ \mathcal{W}))|\Upsilon'\rangle$$

as before. We now bound

$$\begin{aligned} \langle \Gamma_Q |^{G_B G_C} \langle \Upsilon'_1 |^{AB^n C^n} |\Upsilon''\rangle^{GAB^n C^n G_B G_C} \\ = 2^{-nQ} \sum_{m'm} \langle m' |^G \langle m' |^{G_B G_C} \langle \Upsilon'_1 |^G |m\rangle^G |\Upsilon''_m\rangle \\ = 2^{-nQ} \sum_m \langle m |^{G_B G_C} \langle \Upsilon'_1 |^G |\Upsilon''_m\rangle^{AB^n C^n G_B G_C} \\ \geq 1 - \epsilon. \end{aligned} \quad (31)$$

The last line uses the estimate (30). Since the construction in the previous coding theorem guarantees that

$$F(|\Phi_{Q_B}\rangle^{A\hat{A}}, \mathcal{D}_1(\text{Tr}_{C^n} \Upsilon_1'^{AB^n C^n})) \geq 1 - \epsilon, \quad (32)$$

we may then employ Lemma 5 to combine these last two estimates to show that the state

$$\Omega^{G_B G_C A\hat{A}} = \text{Tr}_{C^n} \mathcal{D}_1(\Upsilon_1'^{GAB^n C^n G_B G_C})$$

which results from the protocol satisfies

$$\begin{aligned} F &\equiv F(|\Gamma_Q\rangle^{G_B G_C} |\Phi_{Q_B}\rangle^{A\hat{A}}, \Omega^{G_B G_C A\hat{A}}) \\ &\geq 1 - 3 \left( 1 - F(|\Phi_{Q_B}\rangle^{A\hat{A}}, \Omega^{A\hat{A}}) \right) \\ &\quad - |\Gamma_Q^{G_B G_C} - \Omega^{G_B G_C}|_1 \\ &\geq 1 - 3\epsilon - \sqrt{8\epsilon} \\ &\geq 1 - 6\sqrt{\epsilon}. \end{aligned}$$

The bound on the fidelity in the second line is from (32), while the bound on the trace distance in the next line is by application of (4) to the square-root of the fidelity in (31). This proves the coding theorem. ■

*Proof: (Converse)* Observe that any  $(Q, Q_B, n, \epsilon)$  qq entanglement generation code is able to establish  $\epsilon$ -good uniform common randomness between Alice, Bob and Charlie at rate  $Q$ , in the sense that they generate a triple of random variables  $(M_A, M_B, M_C)$  which satisfy

$$|\text{dist}(M_A, M_B, M_C) - \text{dist}(M, M, M)|_1 \leq 2\epsilon,$$

where  $M$  is uniformly distributed on  $\{1, \dots, 2^{nR}\}$ . To accomplish this, Alice will measure the  $G$  part of her input  $\Upsilon^{GAA^n}$  in the GHZ basis  $\{|m\rangle^G\}$  at any point in the protocol, while Bob and Charlie measure their respective bases  $\{|m\rangle^{G_B}\}$  and  $\{|m\rangle^{G_C}\}$  after their decodings are complete. The previous converse provides an upper bound on this uniform common randomness generation rate for protocols which also generate Alice-Bob entanglement at rate  $Q$ , therefore proving this converse. ■

#### F. Proof of Theorem 6

*Proof: (Converse)* As per the remarks in the previous converse theorem, the converse part of the proof of Theorem 4 applies here as well. ■

## APPENDIX I

### PROOF OF CARDINALITY BOUNDS FOR $\mathcal{T}$

In Theorem 1 (with  $k = 1$ ), let a finite set  $\mathcal{T}$  and conditional probabilities  $p(x|t)$  be arbitrary. Geometrically, this amounts to fixing a  $\mathcal{T}$ -labeled set of points on the  $\mathcal{X}$ -probability simplex. We will show that given any probabilities  $p(t)$  on  $\mathcal{T}$ , there exists another distribution  $q(t)$  which puts positive mass on at most  $\min\{|\mathcal{X}|, |B|^2 + |C|^2 - 1\}$  elements of  $\mathcal{T}$ , while satisfying

$$\begin{aligned} I(X; C|T)_q &= I(X; C|T)_p \\ I(T; B)_q &= I(T; B)_p, \\ I(T; C)_q &= I(T; C)_p, \end{aligned}$$

where the subscript  $q$  means the quantity is evaluated on the state

$$\rho_q^{TXBC} \equiv \bigoplus_{x,t} q(t) p(x|t) \rho_x^{BC}.$$

We will prove this by use of the following lemma:

*Lemma 6 (Fenchel and Eggleston [12]):* Let  $\mathcal{S} \subset \mathbb{R}^n$  have at most  $n$  connected components. Then any point in the convexification of  $\mathcal{S}$  can be written as a convex combination of at most  $n$  points in  $\mathcal{S}$ .

It will thus be sufficient to show that the map

$$f: p(t) \mapsto (I(X; C|T)_p, I(T; B)_p, I(T; C)_p)$$

factors through an affine space of sufficiently low dimension. To this end, we decompose  $f$  into a nonlinear part  $f_{\text{nl}}$  and an affine part

$$\begin{aligned} f_{\text{aff}}: p(t) &\mapsto \sum_t p(t) (I(X; B|t), \rho_t^B, \rho_t^C) \\ &= (I(X; B|T), \rho^B, \rho^C), \end{aligned} \quad (33)$$

so that the following diagram commutes:

$$\begin{array}{ccc} p(t) & \xrightarrow{f} & (I(X; C|T)_p, I(T; B)_p, I(T; C)_p) \\ & \searrow f_{\text{aff}} & \uparrow f_{\text{nl}} \\ & & (I(X; B|T), \rho^B, \rho^C). \end{array}$$

We regard the affine map as producing convex combinations of the points in some affine space parameterizations of the  $\{(I(X; B|t), \rho_t^B, \rho_t^C)\}_{t \in \mathcal{T}}$ , weighted by the probabilities  $p(t)$ . As  $\rho_t^B$  and  $\rho_t^C$  can be specified either by their individual parameterizations or by  $p(x|t)$ , the more efficient representation requires at most  $\min\{|\mathcal{X}| - 1, |B|^2 - 1 + |C|^2 - 1\}$  numbers. Since the first coordinate can be taken to be  $I(X; B|t)$  itself, we see that at most  $\min\{|\mathcal{X}|, |B|^2 + |C|^2 - 1\}$  affine parameters are required to describe (33). By continuity, the image of the  $\mathcal{T}$ -simplex under  $f_{\text{aff}}$  is connected, and so we may use the earlier lemma to infer the existence of probabilities  $q(t)$  on  $\mathcal{T}$  with support cardinality at most  $\min\{|\mathcal{X}|, |B|^2 + |C|^2 - 1\}$ , while satisfying  $f(p(t)) = f(q(t))$ .

For Theorem 2, the degradedness of the channel implies that the  $|C|^2 - 1$  affine parameters of  $\rho^C$  depend affinely on those of  $\rho^B$ , allowing the reduction of the cardinality bound to  $|\mathcal{T}| \leq \min\{|\mathcal{X}|, |B|^2\}$ .



For the bound of Theorem 3, we instead begin by fixing states  $\{\rho_t^{A'}\}_{t \in \mathcal{T}}$ . Here, the affine map outputs convex combinations of the points  $(I_c(\rho_t^{A'}, \mathcal{N}_B), \rho_t^B, \rho_t^C)$ . As a parameterization of the possible  $\rho_t^B$  and  $\rho_t^C$  requires no more than  $\min\{|A'|^2 - 1, |B|^2 - 1 + |C|^2 - 1\}$  coordinates, we obtain by similar reasoning as above that it suffices to take  $|\mathcal{T}| \leq \min\{|A'|^2, |B|^2 + |C|^2 - 1\}$ .

The bound for Theorem 4 follows in the same way as that for Theorem 2, although the fact that  $|B| = |\mathcal{X}|$  implies that  $|\mathcal{T}| \leq |\mathcal{X}|$  is sufficient. In Theorems 5 and 6, the bounds are the same as those from Theorems 3 and 4 and follow for the same reasons.

#### ACKNOWLEDGEMENTS

Thanks! ID was supported by the NSF grant no. CCF-0524811. JY is thankful for support from the Canadian Institute for Advanced Research and the hospitality of the Department of Computer Science at McGill University, where the majority of this research was conducted. He received additional support from the US National Science Foundation under grant EIA-0086038 through the Institute for Quantum Information at the California Institute of Technology, where this manuscript was completed. PH is grateful for support from the Canada Research Chairs program, the Canadian Institute for Advanced Research and NSERC. JY would also like to thank Tom Cover and Young-Han Kim for helpful discussions, as well as Andreas Winter and Aram Harrow for comments on an earlier version of this manuscript.

#### REFERENCES

- [1] H. Barnum, E. Knill, M. A. Nielsen, "On Quantum Fidelities and Channel Capacities," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1317–1329, 2000.
- [2] C.H. Bennett, D. DiVincenzo, J. Smolin, "Capacities of quantum erasure channels," *Phys. Rev. Lett.*, vol. 78, pp. 3217–3220, April 1997.
- [3] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, pp. 197–207, March 1973.
- [4] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2–14, January 1972.
- [5] T. Cover, J. A. Thomas, *Elements of Information Theory*, John-Wiley & Sons, Inc., 1991.
- [6] I. Csiszar, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akadémiai Kiadó, Budapest.
- [7] I. Csiszar, J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, May 1978.
- [8] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inform. Theory*, vol. IT-55, pp. 44–55, January 2005, quant-ph/0304127.
- [9] I. Devetak, P. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," quant-ph/0311131.
- [10] I. Devetak, A. Winter, "Distilling common randomness from bipartite quantum states," quant-ph/0304196.
- [11] I. Devetak, A. Winter, "Distillation of secret key and entanglement from quantum states," *Proc. Roy. Soc. A* 461, 207–235, 2005, quant-ph/0306078.
- [12] H. G. Eggleston, *Convexity*, Cambridge University Press, N.Y. 1963.
- [13] M. Fannes, "A continuity property of the entropy density for spin lattices," *Commun. Math. Phys.* 31, pp. 291–294, 1973.
- [14] C. A. Fuchs, J. van de Graaf, "Cryptographic Distinguishability Measures for Quantum Mechanical States," *IEEE Trans. Inform. Theory*, vol. IT-45, pp. 1216–1227, 1999.
- [15] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 17–25, 1980.
- [16] M. Hayashi, private communication.
- [17] M-H. Hsieh, I. Devetak, A. Winter, "Entanglement-Assisted Capacity of Quantum Multiple Access Channels", quant-ph/0511228.
- [18] P. Hayden, R. Jozsa, D. Petz, A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," *Commun. Math. Phys.* 246(2), pp. 359–374, 2004, quant-ph/0304007.
- [19] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum channel," *Probl. Inf. Transm.*, vol. 9, pp. 177–183, 1973.
- [20] A. S. Holevo, "The capacity of the quantum channel with general input states," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, p. 269, quant-ph/9611023.
- [21] M. Horodecki, J. Oppenheim, A. Winter, "Partial quantum information," *Nature*, quant-ph/0505062.
- [22] M. Horodecki, J. Oppenheim, A. Winter, quant-ph/0512247.
- [23] E. H. Lieb, M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.* 14, pp. 1938–1941, Dec. 1973.
- [24] S. Lloyd, "Capacity of the noisy quantum channel," *Phys. Rev. A*, vol. 55, no. 3, pp. 1613–1622, 1997.
- [25] M. A. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2001.
- [26] B. Schumacher, M. A. Nielsen, "Quantum Data Processing and Error Correction," *Phys. Rev. A* vol. 54, pp. 2629–2635, 1996, quant-ph/9604022.
- [27] B. Schumacher, M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, p. 131.
- [28] P. Shor, "The quantum channel capacity and coherent information," lecture notes, MSRI Workshop on Quantum Computation, 2002.
- [29] J. Smolin, F. Verstraete, A. Winter, "Entanglement of assistance and multipartite state distillation," quant-ph/0505038.
- [30] A. Winter, "The capacity of the quantum multiple access channel," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3059–3065, November 2001, quant-ph/9807019.
- [31] A. Wyner, "The wire-tap channel," *Bell Systems Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [32] J. Yard, I. Devetak, P. Hayden, "Capacity theorems for quantum multiple access channels – classical-quantum and quantum-quantum capacity regions," quant-ph/0501045.